# Use of machine learning models to predict prognosis of combined pulmonary fibrosis and emphysema in a Chinese population

Qing Liu, Di Sun, Yu Wang, Pengfei Li, Tianci Jiang, Lingling Dai, Mengjie Duo, Ruhao Wu and Zhe Cheng[*]

## Abstract

**Background:** Combined pulmonary fibrosis and emphysema (CPFE) is a novel clinical entity with a poor prognosis. This study aimed to develop a clinical nomogram model to predict the 1-, 2- and 3-year mortality of patients with CPFE by using the machine learning approach, and to validate the predictive ability of the interstitial lung disease-gender-age-lung physiology (ILD-GAP) model in CPFE.

**Methods:** The data of CPFE patients from January 2015 to October 2021 who met the inclusion criteria were retrospectively collected. We utilized LASSO regression and multivariable Cox regression analysis to identify the variables associated with the prognosis of CPFE and generate a nomogram. The Harrell's C index, the calibration curve and the area under the receiver operating characteristic (ROC) curve (AUC) were used to evaluate the performance of the nomogram. Then, we performed likelihood ratio test, net reclassification improvement (NRI), integrated discrimination improvement (IDI) and decision curve analysis (DCA) to compare the performance of the nomogram with that of the ILD-GAP model.

**Results:** A total of 184 patients with CPFE were enrolled. During the follow-up, 90 patients died. After screening out, diffusing lung capacity for carbon monoxide (DLCO), right ventricular diameter (RVD), C-reactive protein (CRP), and globulin were found to be associated with the prognosis of CPFE. The nomogram was then developed by incorporating the above five variables, and it showed a good performance, with a Harrell's C index of 0.757 and an AUC of 0.800 (95% CI 0.736–0.863). Moreover, the calibration plot of the nomogram showed good concordance between the prediction probabilities and the actual observations. The nomogram also improved the discrimination ability of the ILD-GAP model compared to that of the ILD-GAP model alone, and this was substantiated by the likelihood ratio test, NRI and IDI. The significant clinical utility of the nomogram was demonstrated by DCA.

**Conclusion:** Age, DLCO, RVD, CRP and globulin were identified as being significantly associated with the prognosis of CPFE in our cohort. The nomogram incorporating the 5 variables showed good performance in predicting the mortality of CPFE. In addition, although the nomogram was superior to the ILD-GAP model in the present cohort, further validation is needed to determine the clinical utility of the nomogram.

**Keywords:** CPFE, Machine learning model, Nomogram, ILD-GAP model, Prognosis

## Introduction

Pulmonary interstitial fibrosis and emphysema are two distinct clinical entities with different pathogeneses and pathophysiologic manifestations. However, an increasing number of studies consider that the two phenotypes

*Correspondence: chengzhezzu@outlook.com

Department of Respiratory and Critical Care Medicine, The First Affiliated Hospital of Zhengzhou University, Zhengzhou, People's Republic of China

Liu *et al. BMC Pulmonary Medicine*      (2022) 22:327

Page 2 of 11

can coexist within one patient [1, 2]. Cottin et al. defined a novel phenotype, "combined pulmonary fibrosis and emphysema (CPFE)", in 2005 [3]. CPFE is a clinical syndrome characterized by the coexistence of emphysema in the upper zones and fibrosis in the bases of the lungs [1, 2]. The median survival time for CPFE patients is reported to be 2.1 to 6.1 years, which is extremely poorer than that of patients with fibrosis or emphysema alone [4, 5]. Therefore, a validated risk assessment is desperately needed for the cognition and management of CPFE patients.

The study of prognostic prediction of CPFE remains challenging because of the heterogeneity in disease-specific variables and the lack of awareness for this clinical entity [5–9]. Unfortunately, research evaluating and establishing a prognostic prediction system for CPFE is rare to date. The interstitial lung disease-gender-age-lung physiology (ILD-GAP) model is widely used to predict the prognosis of chronic ILD subtypes, including idiopathic pulmonary fibrosis (IPF), connective tissue disease associated ILD (CTD-ILD) and unclassifiable ILD [10]. The previous researches indicated that the prevalence of emphysema was around 27% in patients with chronic ILD subtypes, including IPF and CTD-ILD [11]. CPFE is a distinct chronic ILD subtype with special clinical features and a poor prognosis [3]. However, previous studies have not performed CPFE subtyping analysis with the ILD-GAP model. Therefore, there is an urgent need to explore the prognostic factors of CPFE and to assess and improve the ILD-GAP model [11].

In this study, we investigated the prognostic factors of CPFE and established a comprehensive nomogram to predict the mortality of CPFE in a Chinese population. Furthermore, we also evaluated the prognostic predictive performance of the ILD-GAP model in CPFE.

## Methods

### Study population

This study was a retrospective cohort study that included 184 confirmed CPFE patients who were admitted to the First Affiliated Hospital of Zhengzhou University between January 2015 and October 2021.

Patients were diagnosed with CPFE according to the criteria suggested by Cottin et al. [3], namely, the radiographic presence of centrilobular and/or paraseptal emphysema ($\geq 10\%$) in the upper zones and pulmonary fibrosis in the bases of the lungs.

CTD was defined according to the criteria recommended by the American Rheumatism Association and the American College of Rheumatology [12–19], including systemic sclerosis (SSc), rheumatoid arthritis (RA), polymyositis/dermatomyositis (PM/DM), sjogren syndrome (SS), ankylosing spondylitis (AS), systemic lupus erythematosus (SLE), antineutrophil cytoplasmic antibody (ANCA)-associated vasculitis (AAV), mixed connective tissue disease (MCTD), and undifferentiated connective tissue disease (UCTD).

The exclusion criteria were as follows: (1) patients who met the criteria for the diagnosis of CPFE, but CPFE was secondary to other etiologies, including pneumoconiosis (asbestosis or siderosis); (2) patients with incomplete data; and (3) patients younger than 18 years old.

### Ethics issue

The ethical approval of this study was granted by the Ethics Committee of Scientific Research and Clinical Trials of the First Affiliated Hospital of Zhengzhou University (approval number: 2019-KY-116) prior to the data collection. Since the data were deidentified and aggregated, written consent was waived.

### Data collection

Data were collected from electronic medical records at the initial diagnosis. The collected data included demographic characteristics, systematic classification, pulmonary function test results, echocardiography results, high-resolution computed tomography (HRCT) images and laboratory test results.

The demographic characteristics included age, sex, body mass index, smoking history, complications (lung cancer and pulmonary hypertension) and treatment. Pulmonary hypertension (PH) was defined according to the echocardiographic criteria for high probability of PH recommended by the European Society of Cardiology and the European Respiratory Society (ESC/ERS): the peak tricuspid regurgitation velocity (TRV) > 3.4 m/s; or the TRV is 2.9–3.4 m/s within the signs assessing the right ventricular (RV) size, the pressure overload, the pattern of blood flow velocity out of the RV, the diameter of the pulmonary artery and an estimate of right atrial pressure [20].

The systematic classifications included idiopathic CPFE and CTD-CPFE, using the classification criteria of CTD as the inclusion standard.

The variables of the pulmonary function test that we collected included the percentage of the predicted values (%Predicted) for forced expiratory volume in the first second ($FEV_1$), forced vital capacity (FVC), total lung capacity (TLC), peak expiratory flow (PEF), maximal midexpiratory flow rate (MMEF, also known as FEF 25–75), DLCO and the ratio of $FEV_1$ to FVC ($FEV_1$/FVC).

The collected echocardiography data included right atrial area (RAA), RVD (from the right ventricular four-chamber view, the straight line joining the midpoint of the tricuspid valve annulus to the right ventricular apex in end-diastole constituted the RVD) [21, 22], left atrial

Liu *et al. BMC Pulmonary Medicine*     (2022) 22:327

Page 3 of 11

area, left ventricular end diastolic diameter, ascending aortic diameter, aortic annulus diameter, pulmonary artery diameter, pulmonary regurgitant peak velocity, and left ventricular ejection fraction (LVEF).

HRCT scans were examined by two independent chest radiologists, and final conclusions on the findings were reached by consensus. The collected HRCT images included fine reticular opacity, ground-glass opacity, pseudoplaque, flocculent shadow (any area that preferentially attenuates the X-ray beam and therefore appears more opaque than the surrounding area), parenchymal band, honeycomb shadow, traction bronchiectasis, local pleural thickening and mediastinal lymphadenopathy. The detailed description of these images refers to the Fleischner terminology [23].

The collected laboratory examination data included leukocyte count, erythrocyte count, haemoglobin count, platelet count, red cell distribution width, platelet distribution width, aspartate transaminase, alanine aminotransferase, γ glutamyl transferase, alkaline phosphatase, total protein, albumin (ALB), globulin, total bilirubin, direct bilirubin, indirect bilirubin, urea nitrogen, creatinine, uric acid, glomerular filtration rate, total cholesterol, total triglycerides, high-density lipoprotein, low-density lipoprotein, B-type natriuretic peptide, CRP, procalcitonin, erythrocyte sedimentation rate, complement component C3, complement component C4, immunoglobulin A, immunoglobulin M, immunoglobulin G, Krebs von den Lungen-6 (KL-6), partial pressure of carbon dioxide, partial pressure of arterial oxygen, blood oxygen saturation, lactate, alpha-fetoprotein, carcinoembryonic antigen, carbohydrate antigen 125, cytokeratin-19-fragment (CYFRA21-1), neuron-specific enolase, carbohydrate antigen 199, carbohydrate antigen 153, carbohydrate antigen 724 and serum ferritin.

### Follow-up and outcome assessment

The study endpoint was all-cause mortality during follow-up until January 2022. Follow-up information was obtained from patients or their families via telephone interviews.
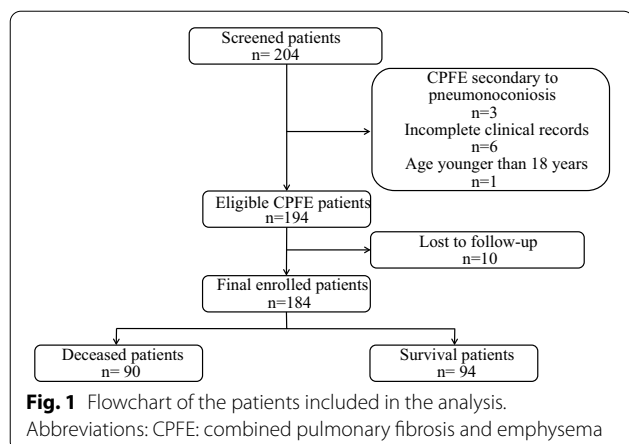
### Statistical analysis

Missing data were processed by multiple imputations. Imputation for missing variables was considered if missing values were less than 20%. A t test or corrected t test was used to compare the continuous variables of normal distribution between the two groups, which were presented as the mean ± standard deviation (mean ± SD). Continuous nonnormally distributed data were compared using the Mann–Whitney U test and presented as the median and interquartile range (IQR, 25–75th percentiles). Categorical variables of the two groups were compared by the $\chi^2$ test and presented as frequencies (percentages). LASSO regression analysis was used for data dimension reduction and variable selection. The penalty value (λ value) was selected by tenfold cross-validation, and the best subset of the variables was selected by using the "glmnet" package of R. The significance of each variable in the best subset was evaluated by univariable Cox regression analysis. The variables with *P* values less than 0.05 were entered into the forward stepwise regression multivariable Cox analysis. A nomogram was constructed based on the results of multivariate Cox regression analysis and by using the "rms" package of R. For clinical use of the model, the risk scores of each patient were calculated based on the nomogram. The performance of the nomogram was assessed by discrimination and calibration [24]. The discriminative ability of the model was determined by the area under the receiver operating characteristic (ROC) curve (AUC). In addition, the nomogram was subjected to 1000 bootstrap resamples for internal validation to assess its predictive accuracy [25]. The calibration of the internal validation model was performed by a visual calibration plot comparing the predicted and actual probability of mortality. The ILD-GAP stage was calculated based on gender (0–1 points), age (0–2 points), and two physiologic lung function parameters—FVC and DLCO (0–5 points) [10]. The predictive performance of the nomogram and ILD-GAP model were evaluated by a likelihood ratio test (using "lmtest" R package), NRI and IDI (using "survC1" and "survIDINRI" R package), the comparison of the Harrell's C index (using "survival" R package) and AUC values (using "ROCR" R package). Finally, decision curve analysis (DCA) was performed by the source file "stdca. R". All analyses were performed using SPSS version 26.0 and R version 4.1.1. For all the analyses, *P* < 0.05 was considered to be statistically significant.

## Results

### Clinical characteristics

A total of 204 patients with confirmed CPFE were screened in the present study according to the above defined criteria. After excluding the patients with CPFE secondary to pneumoconiosis (n = 3), patients with incomplete data (n = 6), those younger than 18 years old (n = 1) and those lost to follow-up (n = 10), 184 patients were included in this study, as presented in Fig. 1. During follow-up (median duration 16.9 months), a total of 90 (48.9%) patients died. The clinical characteristics of all patients in the study are shown in Table 1 (at the end of the article). In our study cohort, 143 (78%) were male, 105 (57%) had a history of smoking, and the mean age at the initial diagnosis was 67 ± 11 years old. The median overall survival time was 32.8 months. Compared with

Liu *et al. BMC Pulmonary Medicine*     (2022) 22:327

Page 4 of 11



**Fig. 1** Flowchart of the patients included in the analysis. Abbreviations: CPFE: combined pulmonary fibrosis and emphysema

the surviving patients, the deceased patients were significantly older, were more likely to have pulmonary hypertension and lung cancer, and treated without acetylcysteine (all $P < 0.05$). Compared with patients who were alive, those who died were more likely to have lower FVC, TLC, DLCO, LVEF and higher RVD (all $P < 0.05$). Additionally, the deceased patients were more likely to have mediastinal lymphadenopathy, higher levels of serum KL-6 and CYFRA21-1 (all $P < 0.05$).

**Model establishment**

Ninety-five prognostic variables were enrolled in this study. First, we reduced the dimension and selected the best prognostic subset of these indicators by LASSO regression analysis. Then, a ten-fold cross validation of the LASSO model was performed for tuning parameter selection via the minimum criteria (Fig. 2A). The track of each prognostic indicator coefficient was observed in the LASSO coefficient profiles with the changing of the log (lambda) in the LASSO algorithm (Fig. 2B). The optimal lambda value was 0.104 (log(lambda): − 2.262) using the LASSO algorithm, and 6 variables were selected as potential influencing factors of prognosis—age, DLCO, RVD, CRP, ALB and globulin. To explore the potential influencing factors associated with the prognosis of CPFE, we further conducted univariate and multivariate Cox regression analyses. Univariable Cox regression analysis revealed that increased age (HR 1.053, 95% CI 1.031–1.076), RVD (HR 1.115, 95% CI 1.074–1.158), CRP (HR 1.009, 95% CI 1.005–1.013) and globulin (HR 1.066, 95% CI 1.038–1.094) were correlated with a higher mortality risk (all $P < 0.001$) (Table 2). However, higher DLCO (HR 0.965, 95% CI 0.952–0.978) and ALB (HR 0.906, 95% CI 0.867–0.948) were correlated with a lower mortality risk (both $P < 0.001$) (Table 2). Significant indicators ($P$ value < 0.05) in the univariate analysis were entered into a multivariate Cox model, and the results showed

that the 5 variables of age, DLCO, RVD, CRP and globulin affected all-cause mortality significantly (all $P < 0.05$) (Table 2). The nomogram for prognostic prediction was established according to the results of the multivariable Cox regression analysis (Fig. 3).

**Performance of the model**

The predictive performance of the nomogram was good in our study cohort. The C-index value was 0.757, and the mean Harrell's C index in the validation cohort constructed by 1000 bootstrap resamples was 0.853. The AUC of the nomogram was 0.800 (95% CI 0.736–0.863) (Fig. 4A). The calibration curves of the nomogram showed high consistency between the predicted and the actual 1-, 2- and 3-year survival probabilities in our study cohort (Fig. 5).

The ILD-GAP model exhibited increasing mortality risk in patents with higher scores by univariate variable Cox regression (HR 1.652, 95% CI 1.391–1.962, $P < 0.001$; Table 2). The C-index value was 0.657 of the ILD-GAP model, which was lower than that of the nomogram (0.757). The likelihood ratio test showed that there was a statistically significant enhancement of the predictive performance when the inclusion of nomogram in the ILD-GAP model ($P < 0.001$), but there was no significant difference when the inclusion of the ILD-GAP model in nomogram ($P = 0.160$) (Table 3). Moreover, the NRIs of the nomogram and the ILD-GAP model for 1-, 2- and 3-year mortality were 0.332 (95% CI 0.086–0.476, $P = 0.013$), 0.362 (95% CI 0.087–0.511, $P = 0.020$) and 0.173 (95% CI − 0.069 to 0.381, $P = 0.120$), respectively, and the IDIs of the nomogram and the ILD-GAP model for 1-, 2- and 3-year mortality were 0.145 (95% CI 0.054–0.213, $P < 0.001$), 0.142 (95% CI 0.057–0.230, $P < 0.001$) and 0.084 (95% CI − 0.030 to 0.175, $P = 0.133$), respectively (Table 4). These results indicated that the nomogram showed a better prognostic performance than the ILD-GAP model in the present cohort. Then, we performed DCA to evaluate the net clinical benefit that the nomogram would bring to patients compared with the ILD-GAP model. In this study, the nomogram showed a better net benefit than the ILD-GAP model for clinical intervention for the optimal decision threshold > 0% (Fig. 6).

**Discussion**

CPFE is a clinical syndrome without full recognition that is characterized by progressive worsening respiratory symptoms and markedly impaired lung diffusion function [2, 3, 26]. Unfortunately, limited studies have reported its prognostic risk factors [7, 27]. Moreover, none of the previous studies developed a prognostic predictive system for CPFE patients. In the present study, we explored the clinical characteristics and prognostic features of CPFE

Liu *et al. BMC Pulmonary Medicine*     (2022) 22:327

Page 5 of 11

**Table 1** Clinical characteristics of the CPFE patients

| Variable | All patients (n = 184) | Deceased patients (n = 90) | Surviving patients (n = 94) | P value* |
|---|---|---|---|---|
| Age, y | 66.8 ± 10.5 | 69.4 ± 9.2 | 64.2 ± 11.0 | 0.001 |
| Sex, n (%) | | | | |
| Male | 143 (77.7) | 71 (78.9) | 72 (76.6) | 0.709 |
| Female | 41 (22.3) | 19 (21.1) | 22 (23.4) | |
| BMI, kg/m$^2$ | 23.6 (21.5, 25.6) | 23.4 (21.4, 25.0) | 23.7 (21.5, 26.0) | 0.379 |
| Smoking, n (%) | 105 (57.1) | 51 (56.7) | 54 (57.5) | 0.915 |
| Comorbidities, n (%) | | | | |
| Pulmonary hypertension | 48 (26.1) | 33 (36.7) | 15 (16.0) | 0.001 |
| Lung cancer | 6 (3.3) | 6 (6.7) | 0 (0.0) | 0.033 |
| Baseline lung function | | | | |
| FEV$_1$, %Predicted | 82.8 (72.4, 90.0) | 82.3 (71.5, 87.9) | 84.0 (73.5, 96.1) | 0.099 |
| FVC, %Predicted | 84.0 (72.4, 91.1) | 82.9 (66.0, 89.1) | 86.1 (74.8, 96.4) | 0.025 |
| FEV$_1$/FVC, % | 77.9 (71.0, 85.3) | 78.0 (72.8, 83.7) | 77.9 (68.5, 85.7) | 0.574 |
| TLC, %Predicted | 74.2 (66.4, 80.3) | 73.1 (63.8, 78.4) | 74.6 (68.5, 83.0) | 0.034 |
| DLCO, %Predicted | 44.0 (36.0, 52.1) | 40.3 (27.0, 47.2) | 46.7 (40.7, 58.6) | < 0.001 |
| Echocardiography | | | | |
| RAA, cm$^2$ | 14.0 ± 3.6 | 14.5 ± 4.5 | 13.7 ± 2.4 | 0.135 |
| RVD, mm | 17.3 ± 3.3 | 18.5 ± 4.0 | 16.1 ± 1.8 | < 0.001 |
| LVEF, % | 63.0 (61.0, 64.0) | 62.0 (60.0, 64.0) | 63.1 (61.8, 65.0) | 0.027 |
| Image of HRCT, n (%) | | | | |
| Fine reticular opacity | 94 (51.1) | 50 (55.6) | 44 (46.8) | 0.235 |
| Ground-glass opacity | 15 (8.2) | 7 (7.8) | 8 (8.5) | 0.856 |
| Patchy shadow | 89 (48.4) | 47 (52.2) | 42 (44.7) | 0.306 |
| Traction bronchiectasis | 16 (8.7) | 6 (6.7) | 10 (10.6) | 0.339 |
| Honeycomb shadow | 43 (23.4) | 23 (25.6) | 20 (21.3) | 0.493 |
| Local pleural thickening | 146 (79.3) | 76 (84.4) | 70 (74.5) | 0.095 |
| Mediastinal lymphadenopathy | 115 (62.5) | 64 (71.1) | 51 (54.3) | 0.018 |
| KL-6, U/ml | 1201.6 (830.0, 1263.8) | 1217.9 (1095.8, 1296.4) | 1175.5 (696.5, 1262.0) | 0.028 |
| CYFRA21-1, ng/ml | 3.8 (2.3, 5.7) | 4.2 (2.7, 6.4) | 3.0 (2.1, 5.3) | 0.015 |
| Treatment, n (%) | | | | |
| Pirfenidone | 76 (41.3) | 36 (40.0) | 40 (42.6) | 0.725 |
| Glucocorticoids | 86 (46.7) | 46 (51.1) | 40 (42.6) | 0.245 |
| Acetylcysteine | 118 (64.1) | 51 (56.7) | 67 (71.3) | 0.039 |
| Immunosuppressive agents | 27 (14.7) | 15 (16.7) | 12 (12.8) | 0.455 |
| Follow-up time, months | 16.9 (6.7, 33.3) | 9.4 (2.3, 21.7) | 21.0 (13.2, 38.6) | < 0.001 |

*CPFE* combined pulmonary fibrosis and emphysema; *mean ± SD* mean ± standard deviation; *BMI* body mass index; *IQR* interquartile range; *FEV1* forced expiratory volume in the first second; *FVC* forced vital capacity; *FEV1/FVC* the ratio of FEV1 to FVC; *TLC* total lung capacity; *DLCO* diffusing lung capacity for carbon monoxide; *RAA* right atrial area; *RVD* right ventricular diameter; *LVEF* left ventricular ejection fraction; *HRCT* high-resolution computed tomography; *KL-6* Krebs von den Lungen-6; *CYFRA21-1* cytokeratin-19-fragment
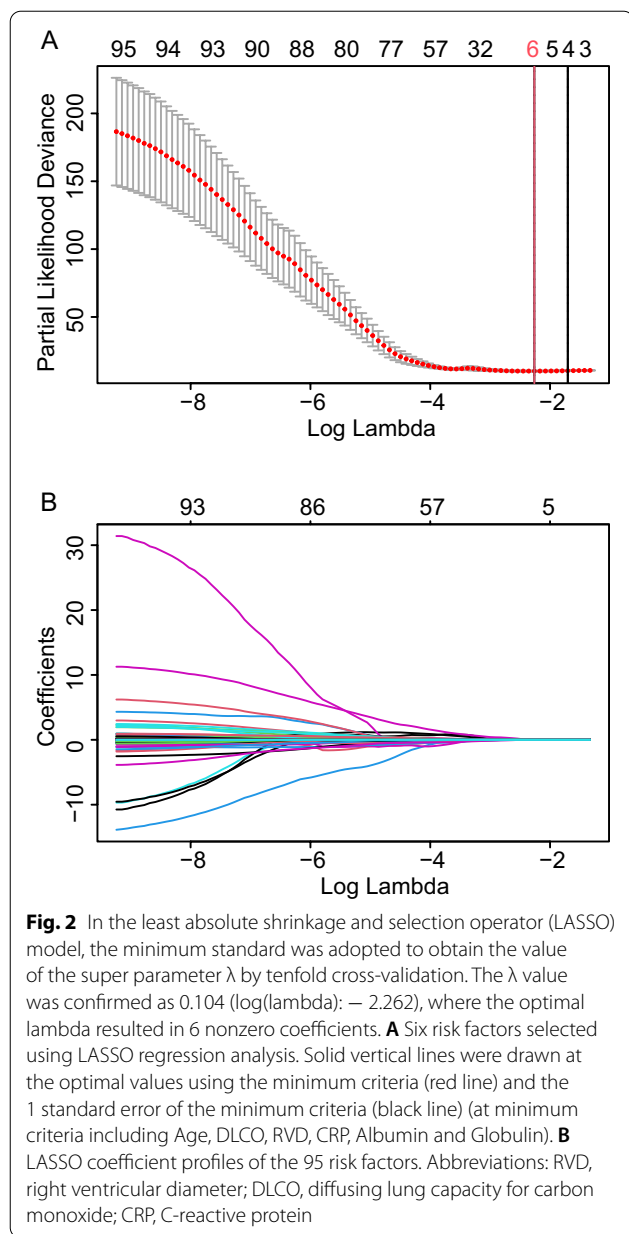
*Comparison of the performance of the deceased patients and the surviving patients for clinical characteristics

Data were presented as the means ± SDs, numbers (%) or medians (IQR)

patients based on the real-world data. Then, we incorporated 5 optimal prognostic variables into a user-friendly nomogram for predicting the prognosis of CPFE. We also performed a series of validations to evaluate the predictive performance of the ILD-GAP model.

We used LASSO regression to screen out 6 variables from the 95 candidates by examining the predictor-outcome association by shrinking the regression coefficients.

LASSO is a method for dimension reduction and variable selection, and the number of selected predictors is not limited by the sample size when the number of samples is more than the number of the variables [28, 29]. There were 95 variables in 180 samples of this study; therefore, LASSO regression was a proper and credible method to establish the model. Then, we used multivariate Cox regression analysis to substantiate the prognostic value of
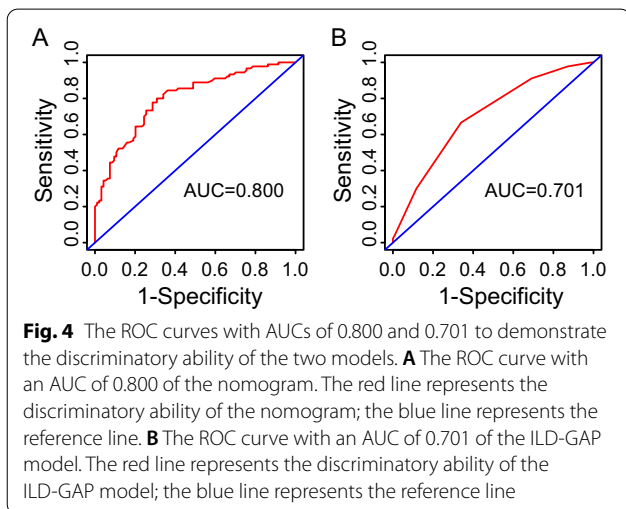
Liu *et al. BMC Pulmonary Medicine*      (2022) 22:327

Page 6 of 11



**Fig. 2** In the least absolute shrinkage and selection operator (LASSO) model, the minimum standard was adopted to obtain the value of the super parameter λ by tenfold cross-validation. The λ value was confirmed as 0.104 (log(lambda): − 2.262), where the optimal lambda resulted in 6 nonzero coefficients. **A** Six risk factors selected using LASSO regression analysis. Solid vertical lines were drawn at the optimal values using the minimum criteria (red line) and the 1 standard error of the minimum criteria (black line) (at minimum criteria including Age, DLCO, RVD, CRP, Albumin and Globulin). **B** LASSO coefficient profiles of the 95 risk factors. Abbreviations: RVD, right ventricular diameter; DLCO, diffusing lung capacity for carbon monoxide; CRP, C-reactive protein
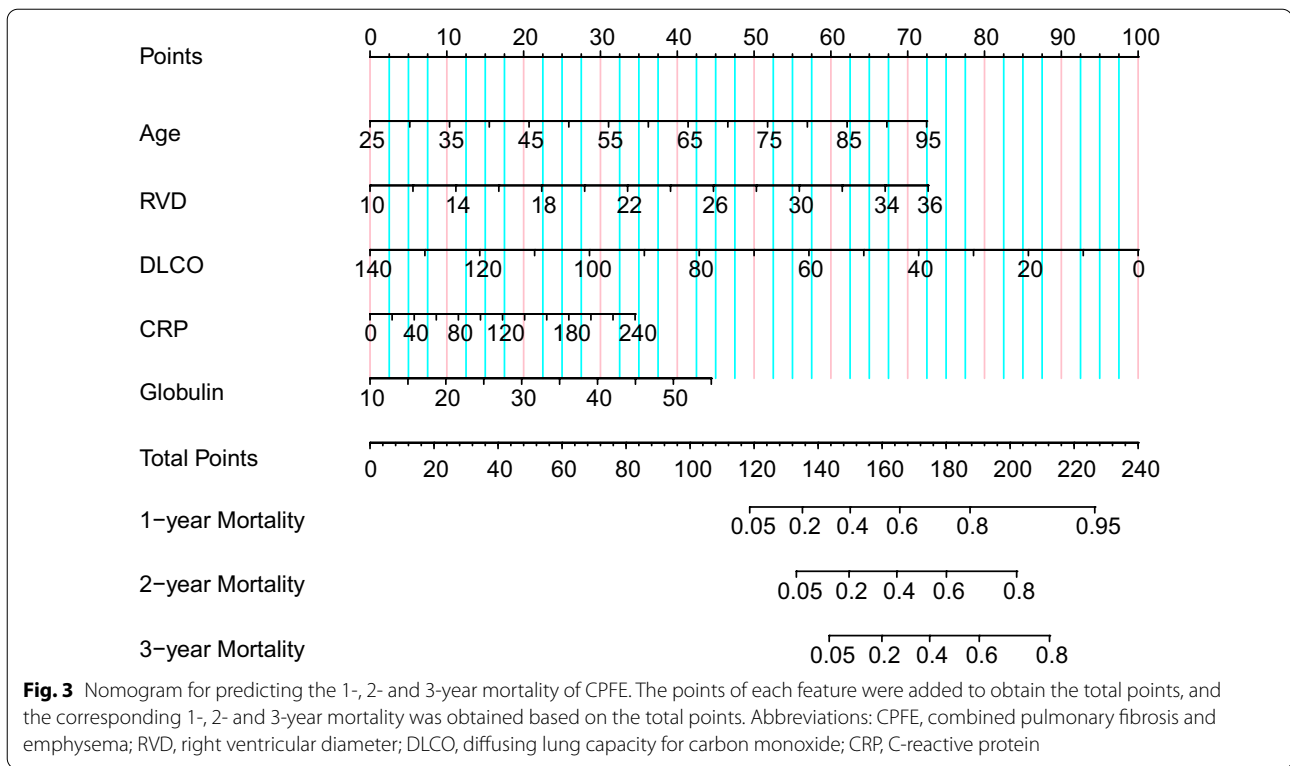
the 6 variables. The results of multivariable Cox analysis showed that age, DLCO, RVD, and the levels of serum CRP and globulin were significantly associated with the prognosis of CPFE.

Age has been reported as a risk factor for the prognosis of CPFE and many other lung diseases because older individuals typically have more comorbidities and poorer health status [30]. CPFE has a mixed pattern of pulmonary function that preserved lung volumes associated with disproportionately reduced DLCO [3]. The cause may be the reduction of normally functioning alveolar capillary units and pulmonary capillary blood volume, which reduces the effective surface area available for gas exchange [2]. In addition, alveolar membrane thickening, excessive accumulation of extracellular matrix and alveolar epithelial cell damage may also be involved in the process of decreasing DLCO according to past studies [2, 31]. Consistently, in this study, a lower DLCO was demonstrated to be associated with the prognosis of CPFE. We also found that RVD was associated with poor prognosis in our study. Ventricular remodelling, mainly showing ventricular wall hypertrophy and cardiac dilatation, is one of the mechanisms of heart failure. Moreover, most patients complicated with heart failure show a reduced cardiac index, which is the most accurate prognostic determinant of CPFE [32]. Previous studies showed that pulmonary hypertension (PH) had a higher prevalence in patients with CPFE with a poor prognosis [33]. In our study, PH was also associated with worse outcomes (HR 2.093, 95% CI 1.360–3.222, $P < 0.05$) in the univariable Cox regression analysis. However, PH was not selected in the LASSO regression analysis. The probable reason is that RVD, which is essentially a marker of right ventricular dilation, may be collinear with PH given that right ventricular dilation is often a consequence of PH. With the progression of PH, the pulmonary artery systolic pressure will decrease with the occurrence of right heart failure, while right ventricular dilation is irreversible.

**Table 2** Univariate and multivariate Cox analyses for overall mortality in CPFE

| Variables | Univariable COX regression analysis | | Multivariable COX regression analysis | |
|---|---|---|---|---|
| | Hazard ratio (95% CI) | *P* value | Hazard ratio (95% CI) | *P* value |
| Age, years | 1.053 (1.031–1.076) | < 0.001 | 1.040 (1.017–1.063) | < 0.001 |
| DLCO, %Predicted | 0.965 (0.952–0.978) | < 0.001 | 0.973 (0.958–0.989) | 0.001 |
| RVD, mm | 1.115 (1.074–1.158) | < 0.001 | 1.111 (1.064–1.161) | < 0.001 |
| CRP, mg/L | 1.009 (1.005–1.013) | < 0.001 | 1.005 (1.001–1.010) | 0.023 |
| Globulin, g/L | 1.066 (1.038–1.094) | < 0.001 | 1.038 (1.005–1.072) | 0.023 |
| ALB, g/L | 0.906 (0.867–0.948) | < 0.001 | | |
| ILD-GAP model | 1.652 (1.391–1.962) | < 0.001 | | |

*CPFE* combined pulmonary fibrosis and emphysema; *DLCO* diffusing lung capacity for carbon monoxide; *RVD* right ventricular diameter; *CRP* C-reactive protein; *ALB* albumin

Liu *et al. BMC Pulmonary Medicine*      (2022) 22:327

Page 7 of 11



**Fig. 3** Nomogram for predicting the 1-, 2- and 3-year mortality of CPFE. The points of each feature were added to obtain the total points, and the corresponding 1-, 2- and 3-year mortality was obtained based on the total points. Abbreviations: CPFE, combined pulmonary fibrosis and emphysema; RVD, right ventricular diameter; DLCO, diffusing lung capacity for carbon monoxide; CRP, C-reactive protein



**Fig. 4** The ROC curves with AUCs of 0.800 and 0.701 to demonstrate the discriminatory ability of the two models. **A** The ROC curve with an AUC of 0.800 of the nomogram. The red line represents the discriminatory ability of the nomogram; the blue line represents the reference line. **B** The ROC curve with an AUC of 0.701 of the ILD-GAP model. The red line represents the discriminatory ability of the ILD-GAP model; the blue line represents the reference line

CRP and globulin are the serological indicators commonly tested in clinic, and their levels are acceptably correlated with the severity of infection and immune status [34]. Our study demonstrated that serum CRP and globulin were significantly associated with increased CPFE mortality. CRP is synthesized in the liver as a result of several stimuli, e.g., interleukin (IL)-6, and is also considered to be a classic acute-phase protein [35, 36]. Several studies have identified that airway damage and acute exacerbations of infection are the main reasons for the poor prognosis for CPFE [2, 27], which could lead to an increase in CRP, consistent with our results. In addition, CPFE could be involved in multiple systems and organs by self-directed inflammation, commonly leading to collagen deposition and tissue damage [37, 38]. CRP is a pattern-recognition molecule of the innate immune system, and its binding to ligands can mediate direct interactions with immunoglobulin receptors and trigger classic complement activation, which is related to the pathogenesis and progression of CPFE [39, 40]. Immune dysregulation is a driver of both idiopathic CPFE and CTD-CPFE [38, 41], and the increasing concentration of immunoglobulin is acceptably related to the active phase and poor prognosis of CPFE [40]. The serum level of globulin could reflect the increase of immunoglobulin and some other abnormal circulating antibodies to some extent.

A nomogram can provide an individualized, evidence-based and highly accurate risk estimation, thus facilitating decision-making by physicians and policy makers [42, 43]. The nomogram we constructed demonstrated good discrimination as assessed by Harrell's C index (0.757) and AUC value (0.800). The optimal calibration curves indicate good consistency between the predicted probabilities and the actual observations, although the variance around the three points shown is high in Fig. 5, which may be due to the relatively small sample size. However,
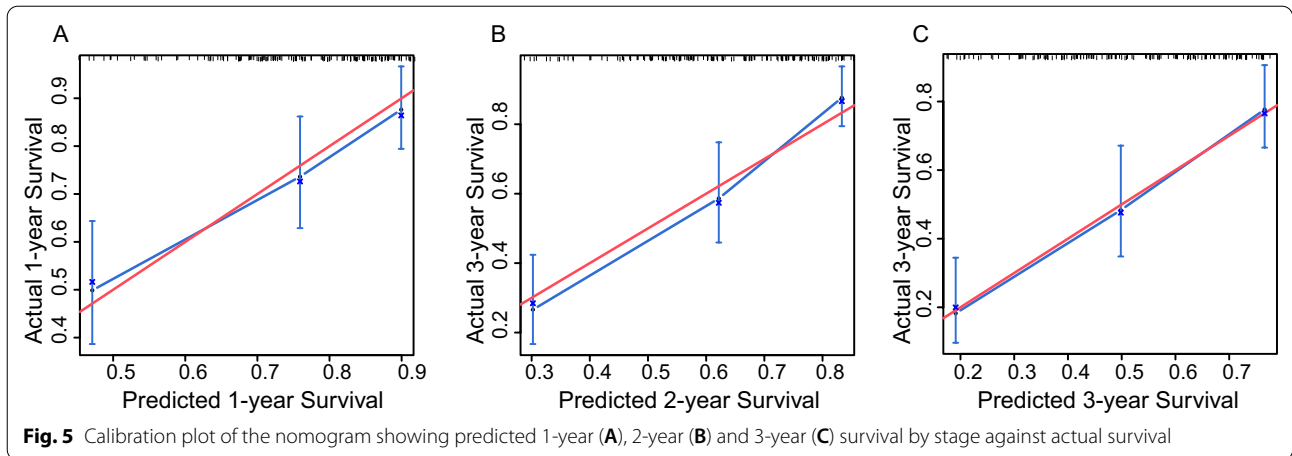
Liu *et al. BMC Pulmonary Medicine*      (2022) 22:327

Page 8 of 11

**Fig. 5** Calibration plot of the nomogram showing predicted 1-year (**A**), 2-year (**B**) and 3-year (**C**) survival by stage against actual survival

**Table 3** Likelihood ratio test between the nomogram and the ILD-GAP model

|  | Nomogram | ILD-GAP model | Nomogram + ILD-GAP model |
|---|---|---|---|
| Likelihood ratio | 78.57 | 33.51 | 80.55 |
| P value | 0.160* | < 0.001# | |

*Comparison of the performance of mortality by nomogram only with that of the combination of the nomogram and the ILD-GAP model

# Comparison of the performance of mortality by the LID-GAP model only with that of the combination of the nomogram and the ILD-GAP model

to our knowledge, our sample size was the largest for the study of the prognosis of CPFE, and the visual representation of the relationship between predicted and observed prognoses was the best way to evaluate calibration [44]. The ILD-GAP model had a good predictive ability in chronic ILD subtypes, but the coexistence of CPFE in IPF/CTD-ILD may affect the existing assessment model [45]. In our cohort, the Harrell's C index of the ILD-GAP model was only 0.657, which indicated poorer discriminative ability than the nomogram (0.757). Moreover, although the combination of the nomogram and the ILD-GAP model was superior to the ILD-GAP model alone, it was not superior to the nomogram alone. The nomogram also improved the discrimination ability compared to the ILD-GAP model substantiated by the NRI and IDI. DCA

has been widely employed to substantiate the clinical utility and benefit when the predictive model guides clinical practice [46]. The DCA proved that when the decision threshold was > 0%, using the nomogram in the current study showed a higher net benefit than using the ILD-GAP model for clinical intervention. We believe this is because of comorbid emphysema, which may impact survival independent of ILD severity [10]. Previous research indicated that $FEV_1$ and $FEV_1/FVC$ were mortality predictors of pulmonary emphysema [47], while there were no significant differences between the deceased and surviving groups. The possible reason may be that we collected the data at the initial diagnosis of CPFE. At the initial stage of the disease, the hyperinflation and high compliance of emphysema probably compensate for the volume loss, which presents relatively preserved lung volume, while even mild emphysema may have an additive effect with fibrosis on the progression of the disease and affect the prognosis of CPFE [3]. The other reason why the nomogram in the current study showed a higher net benefit than the ILD-GAP model may be that there were prognostically important variables for this population that were not captured in the ILD-GAP model. The variables in the nomogram are comprehensive and acquired easily, and can be widely applied to clinical practice after further validation and improvement. Although the

**Table 4** NRI and IDI of the nomogram and the ILD-GAP model in mortality prediction

|  | NRI (95% CI) | P value | IDI (95% CI) | P value |
|---|---|---|---|---|
| 1-Year mortality | 0.332 (0.086–0.476) | 0.013 | 0.145 (0.054–0.213) | < 0.001 |
| 2-Year mortality | 0.362 (0.087–0.511) | 0.020 | 0.142 (0.057–0.230) | < 0.001 |
| 3-Year mortality | 0.173 (− 0.069 to 0.381) | 0.120 | 0.084 (− 0.030 to 0.175) | 0.133 |

*NRI* net reclassification improvement; *IDI* integrated discrimination improvement

Liu *et al. BMC Pulmonary Medicine*    (2022) 22:327
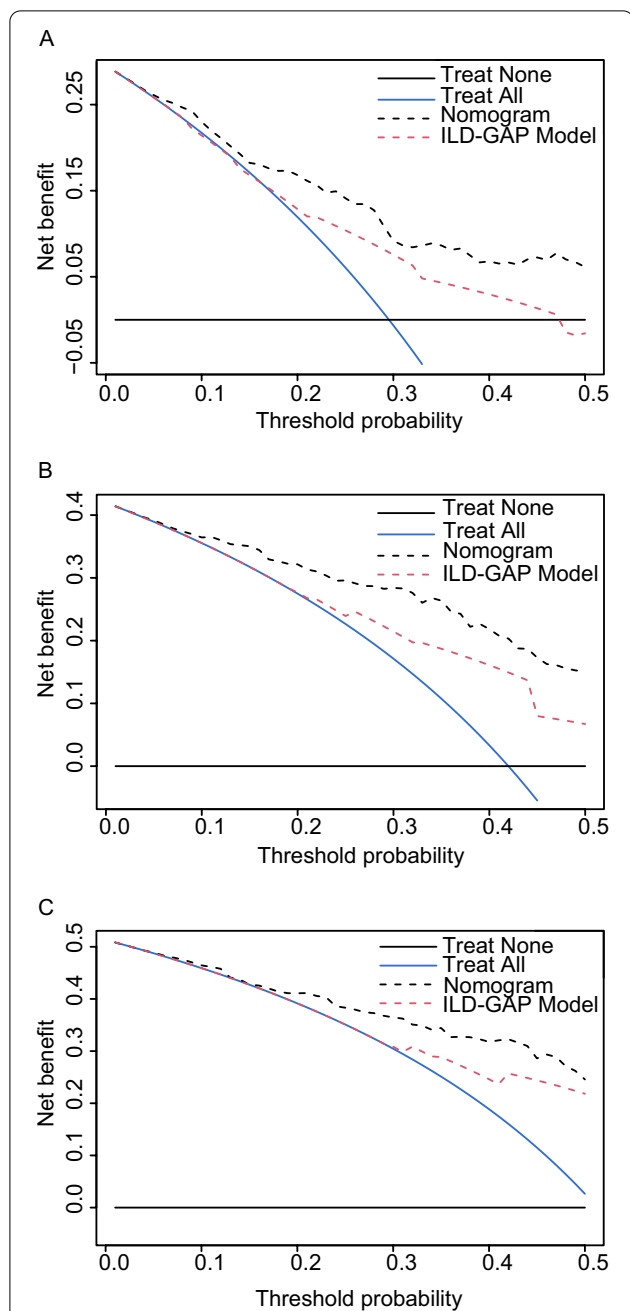
Page 9 of 11



**Fig. 6** Decision curve analysis comparing the clinical performance of the nomogram and ILD-GAP model. For the risk of 1-year (**A**), 2-year (**B**) and 3-year (**C**) mortality, the nomogram showed the highest net benefit for all potential thresholds. The black dotted line represents the nomogram, and the red dotted line represents the ILD-GAP model. The blue line represents the assumption that all patients have been treated, and the black line represents the assumption that no patients have been treated

Although our study is based on real-world data and has relatively complete information of the patients, there are still some limitations. First, this retrospective cross-sectional analysis was based on data from a single institution and therefore may suffer from selection bias. Therefore, more prospective and longitudinal studies are required to further validate the reliability of the nomogram. Second, the nomogram lacked specific genetic markers. However, our study screened 95 clinical characteristics and then selected the most significant variables associated with the prognosis of CPFE. These variables are comprehensive and easily available, thus facilitating decision-making by physicians. Third, quantitative indicators of fibrosis or emphysema were not included in the study, but the more objective indicators, lung function parameters, were included because of the risk of collinearity [48, 49]. Fourth, the diagnosis of PH in the study was based on echocardiography instead of right heart catheterization (RHC). However, a previous study indicated that echocardiography had a specificity of 100% to identify patients with PH and a negative predictive value of 84.72% to rule out PH [50]. Fifth, owing to their small sample sizes, other subtypes of CPFE, such as pneumoconiosis-associated CPFE, were not enrolled to maintain the consistency of baseline data as much as possible. To the best of our knowledge, this is the first model established for predicting the overall mortality of CPFE based on a large sample, and we believe that an early report is urgent and crucial to provide a basis for further studies.

## Conclusion

In this study, age, DLCO, RVD, CRP, and globulin were identified as significant predictive factors of prognosis for CPFE patients. Then, we established a nomogram incorporating the 5 variables to predict the mortality of CPFE, and the nomogram showed good performance. In addition, the nomogram was superior to the ILD-GAP model in terms of performance in the present cohort. Finally, although our nomogram can facilitate individualized therapy design, further validation is still needed to determine the clinical utility of the nomogram.

predictive ability of the ILD-GAP model in CPFE was not superior, the model has been widely adopted to predict the mortality of chronic ILD due to its conciseness and established performance [10].

Liu *et al. BMC Pulmonary Medicine*       (2022) 22:327

Page 10 of 11

## Acknowledgements

## Author contributions
QL, DS, YW, PF-L, TC-J, LL-D, MJ-D and RH-W selected the patients and acquired the data; QL designed the study, analyzed the data and completed the writing. DS and YW were substantially involved in revising the article. ZC had full access to all the data and taken responsibility for the integrity of the data and the accuracy of the data analysis in the study. All authors contributed to the article and approved the submitted version.

## Funding

## Availability of data and materials
All data generated or analyzed during this study are included in this published article. Besides, any additional data/files may be obtained from the corresponding author on reasonable request.

## Declarations

### Ethics approval and consent to participate
The study was performed in accordance with the principles of the Declaration of Helsinki and was approved by the Ethics Committee of Scientific Research and Clinical Trials of the First Affiliated Hospital of Zhengzhou University (approval number: 2019-KY-116). The Ethics Committee of Scientific Research and Clinical Trials of the First Affiliated Hospital of Zhengzhou University granted a waiver of informed consent due to the study's retrospective nature.

### Consent for publication
Not applicable.

### Competing interests
The authors declare that they have no competing interests.

## References
1. Papaioannou AI, Kostikas K, Manali ED, Papadaki G, Roussou A, Kolilekas L, Borie R, Bouros D, Papiris SA. Combined pulmonary fibrosis and emphysema: the many aspects of a cohabitation contract. Respir Med. 2016;117:14–26.
2. Jankowich MD, Rounds SIS. Combined pulmonary fibrosis and emphysema syndrome: a review. Chest. 2012;141(1):222–31.
3. Cottin V, Nunes H, Brillet PY, Delaval P, Devouassoux G, Tillie-Leblond I, Israel-Biet D, Court-Fortune I, Valeyre D, Cordier JF, et al. Combined pulmonary fibrosis and emphysema: a distinct underrecognised entity. Eur Respir J. 2005;26(4):586–93.
4. Lee CH, Kim HJ, Park CM, Lim KY, Lee JY, Kim DJ, Yeon JH, Hwang SS, Kim DK, Lee SM, et al. The impact of combined pulmonary fibrosis and emphysema on mortality. Int J Tuberc Lung Dis. 2011;15(8):1111–6.
5. Jiang CG, Fu Q, Zheng CM. Prognosis of combined pulmonary fibrosis and emphysema: comparison with idiopathic pulmonary fibrosis alone. Ther Adv Respir Dis. 2019;13:1753466619888119.
6. Ryerson CJ, Hartman T, Elicker BM, Ley B, Lee JS, Abbritti M, Jones KD, King TE Jr, Ryu J, Collard HR. Clinical features and outcomes in combined pulmonary fibrosis and emphysema in idiopathic pulmonary fibrosis. Chest. 2013;144(1):234–40.
7. Cottin V. The impact of emphysema in pulmonary fibrosis. Eur Respir Rev. 2013;22(128):153–7.
8. Mitchell PD, Das JP, Murphy DJ, Keane MP, Donnelly SC, Dodd JD, Butler MW. Idiopathic pulmonary fibrosis with emphysema: evidence of synergy among emphysema and idiopathic pulmonary fibrosis in smokers. Respir Care. 2015;60(2):259–68.
9. Jankowich MD, Polsky M, Klein M, Rounds S. Heterogeneity in combined pulmonary fibrosis and emphysema. Respiration. 2008;75(4):411–7.
10. Ryerson CJ, Vittinghoff E, Ley B, Lee JS, Mooney JJ, Jones KD, Elicker BM, Wolters PJ, Koth LL, King TE Jr, et al. Predicting survival across chronic interstitial lung disease: the ILD-GAP model. Chest. 2014;145(4):723–8.
11. Koo BS, Park KY, Lee HJ, Kim HJ, Ahn HS, Yim SY, Jun JB. Effect of combined pulmonary fibrosis and emphysema on patients with connective tissue diseases and systemic sclerosis: a systematic review and meta-analysis. Arthritis Res Ther. 2021;23(1):100.
12. van den Hoogen F, Khanna D, Fransen J, Johnson SR, Baron M, Tyndall A, Matucci-Cerinic M, Naden RP, Medsger TA Jr, Carreira PE, et al. 2013 classification criteria for systemic sclerosis: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis. 2013;72(11):1747–55.
13. Aletaha D, Neogi T, Silman AJ, Funovits J, Felson DT, Bingham CO 3rd, Birnbaum NS, Burmester GR, Bykerk VP, Cohen MD, et al. 2010 rheumatoid arthritis classification criteria: an American College of Rheumatology/European League Against Rheumatism collaborative initiative. Ann Rheum Dis. 2010;69(9):1580–8.
14. Rider LG, Ruperto N, Pistorio A, Erman B, Bayat N, Lachenbruch PA, Rockette H, Feldman BM, Huber AM, Hansen P, et al. 2016 ACR-EULAR adult dermatomyositis and polymyositis and juvenile dermatomyositis response criteria-methodological aspects. Rheumatology (Oxford). 2017;56(11):1884–93.
15. Shiboski CH, Shiboski SC, Seror R, Criswell LA, Labetoulle M, Lietman TM, Rasmussen A, Scofield H, Vitali C, Bowman SJ, et al. 2016 American College of Rheumatology/European League Against Rheumatism classification criteria for primary Sjogren's syndrome: a consensus and data-driven methodology involving three international patient cohorts. Ann Rheum Dis. 2017;76(1):9–16.
16. Calandrino RL, McAuliffe KJ, Dolmage LE, Trivedi ER. Synthesis of the C3 and C1 constitutional isomers of trifluorosubphthalocyanine and their fluorescence within MDA-MB-231 breast tumor cells. Molecules. 2019;24(21):3832.
17. Aringer M, Costenbader K, Daikh D, Brinks R, Mosca M, Ramsey-Goldman R, Smolen JS, Wofsy D, Boumpas DT, Kamen DL, et al. 2019 European League Against Rheumatism/American College of Rheumatology classification criteria for systemic lupus erythematosus. Ann Rheum Dis. 2019;78(9):1151–9.
18. Chung SA, Langford CA, Maz M, Abril A, Gorelik M, Guyatt G, Archer AM, Conn DL, Full KA, Grayson PC, et al. 2021 American College of Rheumatology/Vasculitis Foundation guideline for the management of antineutrophil cytoplasmic antibody-associated vasculitis. Arthritis Rheumatol. 2021;73(8):1366–83.
19. Mosca M, Tani C, Vagnani S, Carli L, Bombardieri S. The diagnosis and classification of undifferentiated connective tissue diseases. J Autoimmun. 2014;48–49:50–2.
20. Galie N, Humbert M, Vachiery JL, Gibbs S, Lang I, Torbicki A, Simonneau G, Peacock A, Vonk Noordegraaf A, Beghetti M, et al. 2015 ESC/ERS Guidelines for the diagnosis and treatment of pulmonary hypertension: the Joint Task Force for the Diagnosis and Treatment of Pulmonary Hypertension of the European Society of Cardiology (ESC) and the European Respiratory Society (ERS): endorsed by: Association for European Paediatric and Congenital Cardiology (AEPC), International Society for Heart and Lung Transplantation (ISHLT). Eur Respir J. 2015;46(4):903–75.
21. Douglas PS, Khandheria B, Stainback RF, Weissman NJ, Brindis RG, Patel MR, Alpert JS, Fitzgerald D, et al. ACCF/ASE/ACEP/ASNC/SCAI/SCCT/SCMR 2007 appropriateness criteria for transthoracic and transesophageal echocardiography: a report of the American College of Cardiology Foundation Quality Strategic Directions Committee Appropriateness Criteria Working Group, American Society of Echocardiography, American

College of Emergency Physicians, American Society of Nuclear Cardiology, Society for Cardiovascular Angiography and Interventions, Society of Cardiovascular Computed Tomography, and the Society for Cardiovascular Magnetic Resonance. Endorsed by the American College of Chest Physicians and the Society of Critical Care Medicine. J Am Soc Echocardiogr. 2007;20(7):787–805.

22. Malik N, Win S, James CA, Kutty S, Mukherjee M, Gilotra NA, Tichnell C, Murray B, Agafonova J, Tandri H, et al. Right ventricular strain predicts structural disease progression in patients with arrhythmogenic right ventricular cardiomyopathy. J Am Heart Assoc. 2020;9(7): e015016.
23. Hansell DM, Bankier AA, MacMahon H, McLoud TC, Muller NL, Remy J. Fleischner Society: glossary of terms for thoracic imaging. Radiology. 2008;246(3):697–722.
24. Pugh SL, Torres-Saavedra PA. Fundamental statistical concepts in clinical trials and diagnostic testing. J Nucl Med. 2021;62(6):757–64.
25. Jin C, Cao J, Cai Y, Wang L, Liu K, Shen W, Hu J. A nomogram for predicting the risk of invasive pulmonary adenocarcinoma for patients with solitary peripheral subsolid nodules. J Thorac Cardiovasc Surg. 2017;153(2):462–9.
26. Kwiatkowska S. IPF and CPFE—the two different entities or two different presentations of the same disease? Adv Respir Med. 2018;86(1):23–6.
27. Zantah M, Dotan Y, Dass C, Zhao H, Marchetti N, Criner GJ. Acute exacerbations of COPD versus IPF in patients with combined pulmonary fibrosis and emphysema. Respir Res. 2020;21(1):164.
28. Tibshirani R. The lasso method for variable selection in the Cox model. Stat Med. 1997;16(4):385–95.
29. Ajana S, Acar N, Bretillon L, Hejblum BP, Jacqmin-Gadda H, Delcourt C, for the BLISAR Study Group. Benefits of dimension reduction in penalized regression methods for high-dimensional grouped data: a case study in low sample size. Bioinformatics. 2019;35(19):3628–34.
30. Kam MLW, Li HH, Tan YH, Low SY. Validation of the ILD-GAP model and a local nomogram in a Singaporean cohort. Respiration. 2019;98(5):383–90.
31. Awano N, Inomata M, Ikushima S, Yamada D, Hotta M, Tsukuda S, Kumasaka T, Takemura T, Eishi Y. Histological analysis of vasculopathy associated with pulmonary hypertension in combined pulmonary fibrosis and emphysema: comparison with idiopathic pulmonary fibrosis or emphysema alone. Histopathology. 2017;70(6):896–905.
32. Seeger W, Adir Y, Barbera JA, Champion H, Coghlan JG, Cottin V, De Marco T, Galie N, Ghio S, Gibbs S, et al. Pulmonary hypertension in chronic lung diseases. J Am Coll Cardiol. 2013;62(25 Suppl):D109-116.
33. Cottin V, Le Pavec J, Prevot G, Mal H, Humbert M, Simonneau G, Cordier JF. Germ"O"P: pulmonary hypertension in patients with combined pulmonary fibrosis and emphysema syndrome. Eur Respir J. 2010;35(1):105–11.
34. Toubi E, Vadasz Z. Innate immune-responses and their role in driving autoimmunity. Autoimmun Rev. 2019;18(3):306–11.
35. Gimeno D, Delclos GL, Ferrie JE, De Vogli R, Elovainio M, Marmot MG, Kivimaki M. Association of CRP and IL-6 with lung function in a middle-aged population initially free from self-reported respiratory problems: the Whitehall II study. Eur J Epidemiol. 2011;26(2):135–44.
36. Del Giudice M, Gangestad SW. Rethinking IL-6 and CRP: why they are more than inflammatory biomarkers, and why it matters. Brain Behav Immun. 2018;70:61–75.
37. Spagnolo P, Distler O, Ryerson CJ, Tzouvelekis A, Lee JS, Bonella F, Bouros D, Hoffmann-Vold AM, Crestani B, Matteson EL. Mechanisms of progressive fibrosis in connective tissue disease (CTD)-associated interstitial lung diseases (ILDs). Ann Rheum Dis. 2021;80(2):143–50.
38. Shenderov K, Collins SL, Powell JD, Horton MR. Immune dysregulation as a driver of idiopathic pulmonary fibrosis. J Clin Investig. 2021;131(2):e143226.
39. Enocsson H, Karlsson J, Li HY, Wu Y, Kushner I, Wettero J, Sjowall C. The complex role of C-reactive protein in systemic lupus erythematosus. J Clin Med. 2021;10(24):5837.
40. Cai R, Wang Q, Zhu G, Zhu L, Tao Z. Increased expression of caspase 1 during active phase of connective tissue disease. PeerJ. 2019;7: e7321.
41. Cottin V, Nunes H, Mouthon L, Gamondes D, Lazor R, Hachulla E, Revel D, Valeyre D, Cordier JF. Groupe d'Etudes et de Recherche sur les Maladies "Orphelines P: combined pulmonary fibrosis and emphysema syndrome in connective tissue disease. Arthritis Rheum. 2011;63(1):295–304.
42. Collins GS, Reitsma JB, Altman DG, Moons KG. Transparent reporting of a multivariable prediction model for individual prognosis or diagnosis (TRIPOD): the TRIPOD statement. BMJ. 2015;350: g7594.
43. Park SY. Nomogram: an analogue tool to deliver digital knowledge. J Thorac Cardiovasc Surg. 2018;155(4):1793.
44. Alba AC, Agoritsas T, Walsh M, Hanna S, Iorio A, Devereaux PJ, McGinn T, Guyatt G. Discrimination and calibration of clinical prediction models: users' guides to the medical literature. JAMA. 2017;318(14):1377–84.
45. Lee SH, Park JS, Kim SY, Kim DS, Kim YW, Chung MP, Uh ST, Park CS, Park SW, Jeong SH, et al. Comparison of CPI and GAP models in patients with idiopathic pulmonary fibrosis: a nationwide cohort study. Sci Rep. 2018;8(1):4784.
46. Vickers AJ, Cronin AM, Elkin EB, Gonen M. Extensions to decision curve analysis, a novel method for evaluating diagnostic tests, prediction models and molecular markers. BMC Med Inform Decis Mak. 2008;8:53.
47. Timmins SC, Diba C, Farrow CE, Schoeffel RE, Berend N, Salome CM, King GG. The relationship between airflow obstruction, emphysema extent, and small airways function in COPD. Chest. 2012;142(2):312–9.
48. Suzuki M, Kawata N, Abe M, Yokota H, Anazawa R, Matsuura Y, Ikari J, Matsuoka S, Tsushima K, Tatsumi K. Objective quantitative multidetector computed tomography assessments in patients with combined pulmonary fibrosis with emphysema: relationship with pulmonary function and clinical events. PLoS ONE. 2020;15(9): e0239066.
49. Feldhaus FW, Theilig DC, Hubner RH, Kuhnigk JM, Neumann K, Doellinger F. Quantitative CT analysis in patients with pulmonary emphysema: is lung function influenced by concomitant unspecific pulmonary fibrosis? Int J Chron Obstruct Pulmon Dis. 2019;14:1583–93.
50. Hammerstingl C, Schueler R, Bors L, Momcilovic D, Pabst S, Nickenig G, Skowasch D. Diagnostic value of echocardiography in the diagnosis of pulmonary hypertension. PLoS ONE. 2012;7(6): e38519.

## Publisher's Note